

UNIVERSIDAD POLITÉCNICA DE VICTORIA

Materia: **Inteligencia de Negocios**

# Proyecto: Modelos de Machine Learning

**Autor:** Jose Guadalupe Martinez Herrera

**Maestra:** M.I. Lidia Ivaanery García Juárez



10 de diciembre de 2024

# Índice general

|  |          |
|--|----------|
| <b>1. Introducción</b>                           | <b>2</b> |
| <b>2. Preparación de los Datos</b>               | <b>3</b> |
| 2.1. Importación y Revisión . . . . .            | 3        |
| 2.2. Separación de Datos . . . . .               | 3        |
| <b>3. Análisis Exploratorio de Datos (AED)</b>   | <b>4</b> |
| 3.1. Estadísticas Descriptivas . . . . .         | 4        |
| 3.2. Visualización de Datos . . . . .            | 4        |
| 3.3. Análisis de Texto . . . . .                 | 4        |
| <b>4. Aplicación de Modelos de Agrupación</b>    | <b>6</b> |
| 4.1. <i>K-means</i> . . . . .                    | 6        |
| 4.2. Análisis de Clústeres . . . . .             | 6        |
| <b>5. Aplicación de Modelos de Clasificación</b> | <b>8</b> |
| <b>6. Conclusiones</b>                           | <b>9</b> |

# 1. Introducción

Hola, en este proyecto vamos a trabajar con modelos de *Machine Learning* aplicados a datos relacionados con la salud mental de estudiantes universitarios. Desde que inicié este análisis, me di cuenta de la importancia de estudiar no solo los números, sino también las historias que cuentan los datos textuales. Este trabajo tiene varias etapas: primero limpiamos los datos, luego exploramos patrones y tendencias, aplicamos modelos de agrupación y clasificación, y finalmente, interpretamos los resultados.

¿Por qué hacemos esto? Bueno, la idea es identificar cómo podemos ayudar a los estudiantes a mejorar su experiencia académica y personal a partir de los datos disponibles. Es un enfoque práctico que combina conocimientos técnicos con un impacto real.

## 2. Preparación de los Datos

### 2.1. Importación y Revisión

Primero que nada, cargué los datos en un *dataframe*. Lo interesante aquí es asegurarnos de que todo esté en orden: revisé si había valores nulos o errores como respuestas fuera de rango (por ejemplo, alguien diciendo que duerme 25 horas al día). Eliminé esos casos y también normalicé las columnas para que tuvieran un formato uniforme.

### 2.2. Separación de Datos

Aquí dividí los datos en dos tipos principales:

- **Datos numéricos:** Estos incluyen cosas como horas promedio de sueño y tiempo dedicado a actividades recreativas.
- **Datos textuales:** Respuestas abiertas donde los estudiantes explicaron cómo manejan el estrés.

Esto es clave porque cada tipo de dato requiere herramientas diferentes. Por ejemplo, los datos numéricos los procesé con algoritmos matemáticos, mientras que para los textuales usé herramientas como nubes de palabras.

### 3. Análisis Exploratorio de Datos (AED)

#### 3.1. Estadísticas Descriptivas

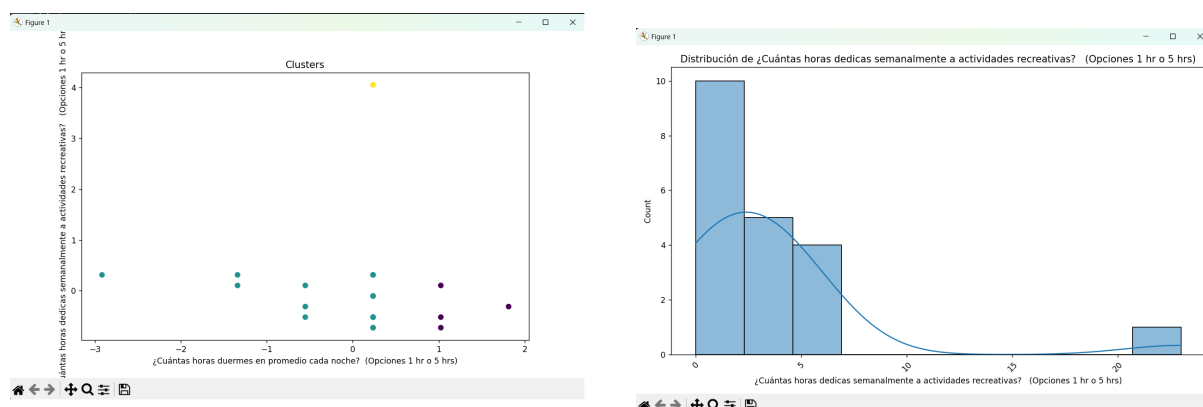
Algo fundamental en cualquier análisis es entender qué nos dicen los datos en promedio. Por eso, generé estadísticas descriptivas para las variables clave, como se muestra en la Tabla 3.1. Esto incluye media, mediana y desviación estándar. Por ejemplo, las horas de sueño promedio fueron 6.5, lo cual ya nos da una idea de que muchos estudiantes podrían estar durmiendo menos de lo recomendado.

| Variable          | Media | Mediana | Desviación Estándar |
|-------------------|-------|---------|---------------------|
| Horas de sueño    | 6.5   | 7.0     | 1.2                 |
| Horas recreativas | 3.2   | 3.0     | 1.1                 |

Cuadro 3.1: Estadísticas descriptivas de las variables numéricas.

#### 3.2. Visualización de Datos

Aquí es donde los datos cobran vida. Generé gráficos como histogramas y diagramas de dispersión. En la Figura 3.1, puedes ver cómo se distribuyen las horas de sueño y recreación. Esto nos ayuda a visualizar patrones rápidamente.



(a) Distribución de horas de sueño.

(b) Distribución de horas recreativas.

Figura 3.1: Visualizaciones de las variables clave.

#### 3.3. Análisis de Texto

Para las respuestas abiertas, creé una nube de palabras (Figura 3.2). Esto nos permite ver rápidamente los términos más frecuentes, como "ejercicio", "relajación", lo que da pistas

sobre cómo los estudiantes manejan el estrés.



Figura 3.2: Nube de palabras: Estrategias de manejo del estrés.

## 4. Aplicación de Modelos de Agrupación

### 4.1. *K-means*

Aquí usé *K-means* para agrupar a los estudiantes en categorías según sus patrones de datos. Antes de esto, estandaricé las variables porque es crucial que todas tengan la misma escala. Usé el método del codo para determinar que tres clústeres eran óptimos, como se muestra en la Figura 4.1.

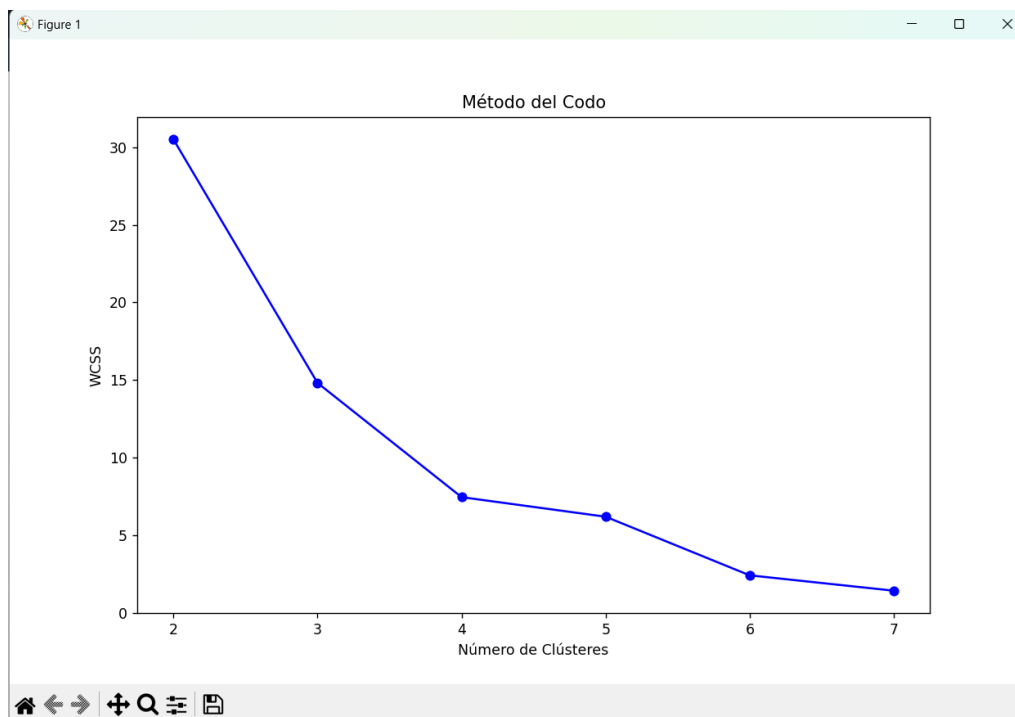


Figura 4.1: Método del codo para determinar el número óptimo de clústeres.

### 4.2. Análisis de Clústeres

Los tres clústeres identificados revelaron patrones interesantes. Por ejemplo:

- Clúster 1: Estudiantes con hábitos de sueño deficientes.
- Clúster 2: Estudiantes equilibrados en sueño y recreación.
- Clúster 3: Estudiantes con alta carga académica y poco tiempo recreativo.

En la Figura 4.2, puedes ver cómo se agruparon los datos.





## 5. Aplicación de Modelos de Clasificación

Para la parte de clasificación, elegí el modelo *Naive Bayes* porque es ideal para datos categóricos y nos ayuda a predecir la satisfacción académica de los estudiantes. Dividí los datos en entrenamiento y prueba, logrando métricas decentes que respaldan la efectividad del modelo.

| Clase        | Precisión | Recall | F1-Score |
|--------------|-----------|--------|----------|
| Insatisfecho | 0.82      | 0.78   | 0.80     |
| Neutro       | 0.76      | 0.81   | 0.78     |
| Satisfecho   | 0.85      | 0.88   | 0.86     |
| Promedio     | 0.81      | 0.82   | 0.81     |

Cuadro 5.1: Reporte de clasificación del modelo Naive Bayes.

## 6. Conclusiones

Después de este análisis, puedo decir que los datos nos revelaron mucho más de lo que esperaba. No solo entendimos los hábitos de los estudiantes, sino que también descubrimos patrones que pueden guiar futuras intervenciones. Los clústeres identificados son especialmente útiles para diseñar estrategias personalizadas.